

OCT 02 2006

Attorney Docket No. 00280758AA

Application Serial No.: 10/749,518

Amendments to the Specification:

Please amend the paragraph beginning at page 6, line 17, and ending at page 6, line 24, as follows:

The second aspect in the process implemented on the data processing system of Figure 1 is that of applying selective sampling based on uncertainty of prediction to enhance the predictive accuracy and to reduce the amount of data required for data analysis. This is done by iteratively selecting data that are "hard to classify" to train the classifier. Then, all classifiers are aggregated to make a final prediction. This process has the effect of weeding out synthesized negative positive examples that are "unrealistic" and too easy to classify (and which do not help in identification of outliers in future test data).

Please amend the paragraph beginning at page 7, line 1 and ending at page 7, line 23, of the specification as follows:

The process implemented by the data processing system shown in Figure 1 is shown in the flow chart of Figure 2. The process begins in step 21 where real ("normal") data is read and stored as T_{real} . In step 22, synthesized ("abnormal") data is generated and stored as T_{syn} . The outlier detection begins at step 23 for Learner A , Data T_{real} , and count t , where $\text{Data } T = T_{\text{real}} \cup T_{\text{syn}}$. Step 24 is a decision block which controls the iteration from $i = 1$ to t . If i is not equal to t , then a determination is made in a second decision block at step 25 to determine if there is more data in T' . If not, selective sampling of data in data storage modules 51 and 52 is done in steps 26 and 27, respectively, to generate selectively sampled data T' , where T' is selectively sampled from data T with sampling probability equaling a measure of uncertainty of prediction for the example or, alternatively, sampling probability proportional to the product of a measure of uncertainty and a measure of cost of mis-classifying the same example. In this process of selective sampling, the given data is scanned and, as the data is scanned, the decision of whether to accept each datum is made independently and with acceptance probability proportional to a

Attorney Docket No. 00280758AA

Application Serial No.: 10/749,518

measure of uncertainty associated with that example. Only those data that are accepted are stored. The selectively sampled and stored data is returned to step 25, after which, in step 28 27, the learning algorithm A is run on data T' to obtain model h_i (i.e., $h_i = A(T')$). The index i of the iteration is incremented by one, i.e., $i = i + 1$, and a return is made to decision block 24. When the iteration is complete, the final model is output at step 27 as $h(x) = \text{sign} \left(\sum_{i=1}^t h_i(x) - \theta \cdot t \right)$.